

文章编号:1001-7372(2007)04-0118-04

## 基于 ARIMA 与人工神经网络组合模型的交通流预测

谭满春<sup>1</sup>, 冯萃斌<sup>1</sup>, 徐建闽<sup>2</sup>

(1. 暨南大学 信息科学技术学院, 广东 广州 510632; 2. 华南理工大学 交通学院, 广东 广州 510641)

**摘要:**将自回归求和滑动平均 (ARIMA) 与人工神经网络组合模型用于短时交通流预测。利用 ARIMA 模型良好的线性拟合能力和人工神经网络强大的非线性关系映射能力, 把交通流时间序列看成由线性自相关结构和非线性结构两部分组成, 采用 ARIMA 模型对交通流序列的线性部分进行预测, 用人工神经网络模型对其非线性残差部分进行预测。结果表明: 组合模型的预测准确性高于各自单独使用时的准确性; 组合方法发挥了 2 种模型各自的优势, 是短期交通流预测的有效方法。

**关键词:** 交通工程; 短期交通流预测; 自回归求和滑动平均模型; 人工神经网络; 时间序列

**中图分类号:** U491.14

**文献标志码:** A

### Traffic Flow Prediction Based on Hybrid ARIMA and ANN Model

TAN Man-chun<sup>1</sup>, FENG Luo-bin<sup>1</sup>, XU Jian-min<sup>2</sup>

(1. School of Information Science and Technology, Jinan University, Guangzhou 510632, Guangdong, China;

2. School of Transportation, South China University of Technology, Guangzhou  
510641, Guangdong, China)

**Abstract:** Hybrid autoregressive integrated moving average (ARIMA) and artificial neural network models were employed in the short-term traffic flow prediction. Using the good linear fitting ability of ARIMA and the strong nonlinear mapping ability of artificial neural network, the traffic flow time series was considered to be composed of a linear autocorrelation structure and a nonlinear structure. ARIMA model was used to predict the linear component of traffic flow time series and the artificial neural network model was applied to the nonlinear residual component prediction. Results show that the hybrid model, which takes advantage of the unique strength of the two models in linear and nonlinear modeling, can produce more accurate predictions than that of single model. The hybrid model can be an efficient method to the short-term traffic flow prediction.

**Key words:** traffic engineering; short-term traffic flow prediction; ARIMA model; artificial neural network; time series

## 0 引言

交通流预测是智能运输系统中的路径诱导和交

通流控制的必要条件, 交通流信息的实时性和可靠性直接关系到交通管理与控制的效果, 因此短时交通流预测是一个研究热点。交通流预测模型与方法

收稿日期: 2006-10-24

基金项目: 国家自然科学基金项目(50578064); 广东省自然科学基金项目(06025219)

作者简介: 谭满春(1968-), 男, 湖南汝城人, 副教授, 工学博士, E-mail: tanmc@jnu.edu.cn。



主要分2类<sup>[1]</sup>:①统计预测算法模型,如简单移动平均、线性回归、自回归滑动平均、Kalman滤波以及非参数回归模型等<sup>[2-6]</sup>;②人工智能或神经网络模型<sup>[7-8]</sup>。研究表明:没有哪一种方法能够适用于所有时间序列的预测,而应当根据实际情况,选择适当的模型与方法<sup>[9-11]</sup>。为了有效地利用各种模型的优点,Bates等<sup>[12]</sup>在1969年提出了组合预测的思想,就是将参与组合的各种预测方法的结果通过适当方式进行组合,扬长避短以获得最优预测结果(至少精度高于各单项方法)。Zhang<sup>[13]</sup>将组合预测模型用于太阳黑子的预测问题,取得了较好的预测效果。

目前通过组合模型进行交通流预测的研究还不多见。本文中笔者利用自回归求和滑动平均(ARIMA)与人工神经网络的组合模型对主干道路的交通流进行短期预测,将统计方法与人工智能方法有机结合,以充分发挥2种模型的优势,从而提高预测准确度。

## 1 交通流预测模型

### 1.1 ARIMA模型

ARIMA模型是由Box等<sup>[14]</sup>提出的一种时间序列建模方法,其建模的基本思想是对非平稳的时间序列用若干次差分使其成为平稳序列,作差分的次数就是参数 $d$ ,再用以 $p, q$ 为参数的ARIMA模型对该平稳序列建模,之后经反变换得到原序列。以 $p, d, q$ 为参数的ARIMA模型预测方程可以表示为

$$y_t = \theta_0 + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \quad (1)$$

式中: $y_t$ 为样本值; $\varphi_i (i=1, 2, \dots, p)$ 和 $\theta_j (j=1, 2, \dots, q)$ 为模型参数; $\varepsilon_t$ 为服从独立正态分布 $N(0, \sigma_a^2)$ 的白噪声序列。 $p, d, q$ 为模型的阶数,通过对不同的 $p, d, q$ 组合测试可以优化模型预测结果,找到最合适的模型参数。

ARIMA建模与预测共有4个步骤。

步骤1:对数据进行平稳化处理。如果数据序列是非平稳的,如存在一定的增长或下降趋势等,则需对数据进行差分处理。

步骤2:模型识别。主要通过分析自回归系数、偏回归系数和互相关系数,为目标序列定阶。

步骤3:参数估计和模型诊断。

步骤4:利用具有合适参数的模型进行预测。

### 1.2 人工神经网络(ANN)预测模型

迄今已有多种不同的神经网络模型被应用于短

期交通流量的预测,例如反向传播(Back Propagation, BP)神经网络、径向基函数神经网络、模糊神经网络等。

BP神经网络模型具有高度非线性关系的映射能力,通过信号的前向传播实现由输入到输出的非线性映射,借助于误差信号的反向传播来对权值不断进行修正。在一个神经网络模型中,除去输入、输出变量外,最核心的变量就是各层之间的连接权重。连接权重通过误差反传算法来确定,模型能够拟合给定的具有非线性关系的历史数据,通过大量历史数据确定连接权重。当一个神经网络模型“学习”得到输入和输出变量的关系之后,即可用于对给定输入的预测。

### 1.3 ARIMA与ANN组合模型

把一组时间序列的数据 $y_t$ 看成是由线性自相关结构 $L_t$ 与非线性结构 $N_t$ 两部分组成的,即

$$y_t = L_t + N_t \quad (2)$$

ARIMA与ANN组合模型预测共有3个步骤。

步骤1:用ARIMA模型对 $y_t$ 进行预测。设预测结果为 $\hat{L}_t$ ,原序列与ARIMA模型预测结果的残差为 $e_t$ ,即

$$e_t = y_t - \hat{L}_t \quad (3)$$

序列 $\{e_t\}$ 中隐含了原序列中的非线性关系

$$e_t = f(e_{t-1}, e_{t-2}, \dots, e_{t-n}) + \varepsilon'_t \quad (4)$$

式中: $\varepsilon'_t$ 为随机误差。

步骤2:用ANN模型对 $\{e_t\}$ 进行预测,即用ANN来逼近非线性函数 $f$ ,并设预测结果为 $\hat{N}_t$ 。

步骤3:用2种模型进行组合预测,结果为

$$\hat{y}_t = \hat{L}_t + \hat{N}_t \quad (5)$$

这就是ARIMA与ANN组合模型的最终预测结果。

从预测过程看,ARIMA模型用于线性部分的预测,而ANN模型用于非线性部分预测,通过对ARIMA和ANN模型的综合运用,发挥了这2种模型的长处,从而提高了预测准确度。

## 2 模型应用

广州市黄埔区夏园观测站位于G107国道广深段(中山大道延长线)上,在该观测站每天24h记录经过该站的各种车辆(如货车、客车、集装箱车等),每小时汇总一次,得到各整点的交通流量序列。选择2005年6月3日至2005年6月22日的数据作为样本,对2005年6月23日的车流量进行预测。

方法1:单独使用ARIMA模型进行预测。根



据 ARIMA 模型的思想,因为原数据是非平稳的序列,所以对原数据进行了 2 次差分变换,把原数据变成平稳的时间序列。通过观察序列的自相关函数和偏相关函数图,初步确定模型参数为:  $p=0\sim 4, d=0\sim 2, q=0\sim 4$ 。经配对试验后,得到 ARIMA 模型,这一过程用 SPSS 软件中的 ARIMA 子程序来实现。由于 ARIMA 模型只适合于进行短期预测,而本文中使用交通流量序列的连续 240 个数据来预测下一整点数据,所以在对 2005 年 6 月 23 日的交通流进行预测的过程中,需要进行 24 次预测得到当天各个整点时刻的预测值。

方法 2:单独使用 ANN 模型进行预测。用 BP 神经网络对上述相同序列进行预测,把 2005 年 6 月 3 日至 2005 年 6 月 20 日的 432 个数据作为训练集,2005 年 6 月 21 日至 22 日的 48 个数据作为测试集,对 2005 年 6 月 23 日的交通流量进行预测。BP 网络有 5 个输入,1 个输出,即用连续 5 项时间序列预测第 6 项。再经过试验,采用双隐层的 BP 神经网络( $5\times 6\times 4\times 1$ ),即双隐层的神经元个数分别是 6 和 4。

方法 3:使用组合模型进行预测,步骤同第 1.3 节,其中的步骤 2 与方法 2 中不同,这一步中使用的训练集和测试集都是来源于序列  $\{e_t\}$ ,见式(3)。

采用 3 个指标来衡量各种方法的预测效果。

(1)百分比绝对误差

$$E_{PAE}(t) = \frac{|y_t - \hat{y}_t|}{y_t} \times 100$$

(2)平均百分比绝对误差

$$E_{MPAE} = \frac{1}{n} \sum_{t=1}^n E_{PAE}(t)$$

(3)相关系数

$$r^2 = \frac{n \sum_{t=1}^n y_t \hat{y}_t - \sum_{t=1}^n y_t \sum_{t=1}^n \hat{y}_t}{\sqrt{(n \sum_{t=1}^n y_t^2 - \sum_{t=1}^n y_t^2)(n \sum_{t=1}^n \hat{y}_t^2 - \sum_{t=1}^n \hat{y}_t^2)}}$$

式中:  $n$  为样本总数。

用以上 3 种方法(ARIMA、ANN 和组合模型)得到的预测结果与实际观测数据的对比见图 1,3 种方法的百分比绝对误差见图 2,2 种指标的对比见表 1。

由图 1、2 和表 1 可见:3 种预测方法中组合模型的  $E_{MPAE}$  最小、相关系数  $r^2$  最大,这表明组合模型的预测效果最好,ANN 模型次之;对于非平稳、随机性较强的交通流序列,ARIMA 模型的预测效果较差;使用 ARIMA 模型只适合进行短期(对本例来说就是 1 h 或 2 h 内)的预测,若进行长期预测则性

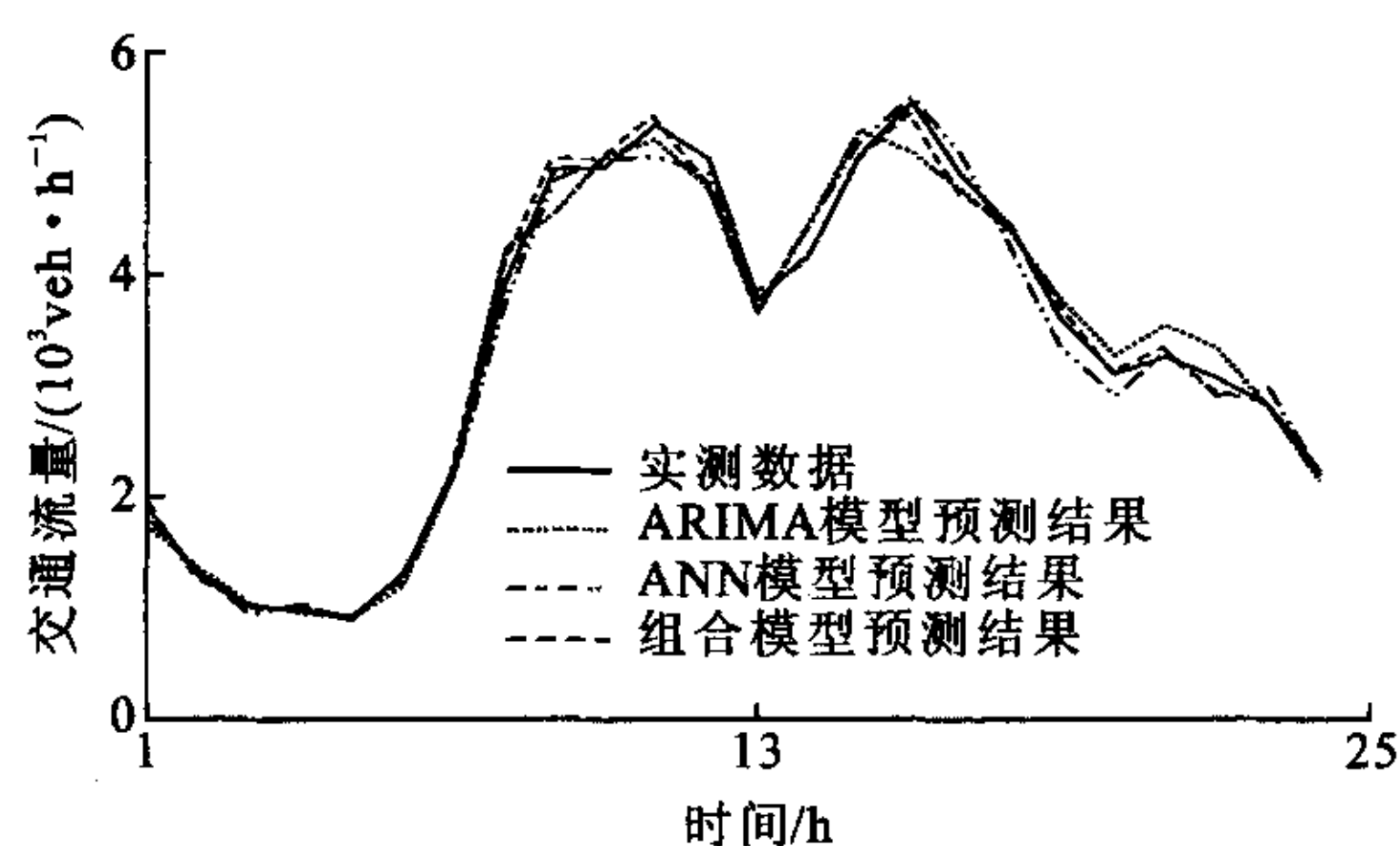


图 1 3 种方法预测结果与实测数据的比较

Fig. 1 Comparison of Field Observed Data and Prediction Results by Three Approaches

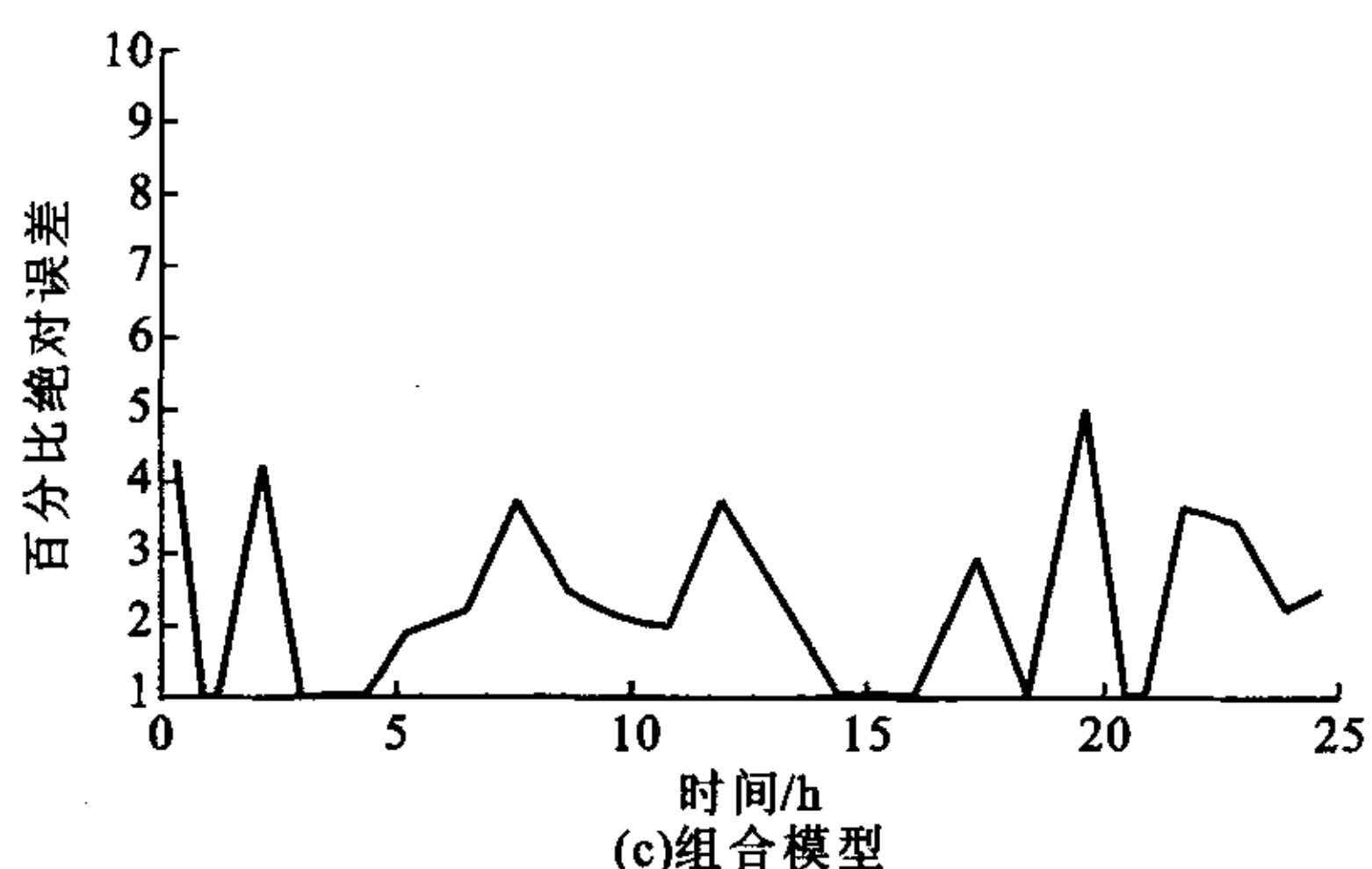
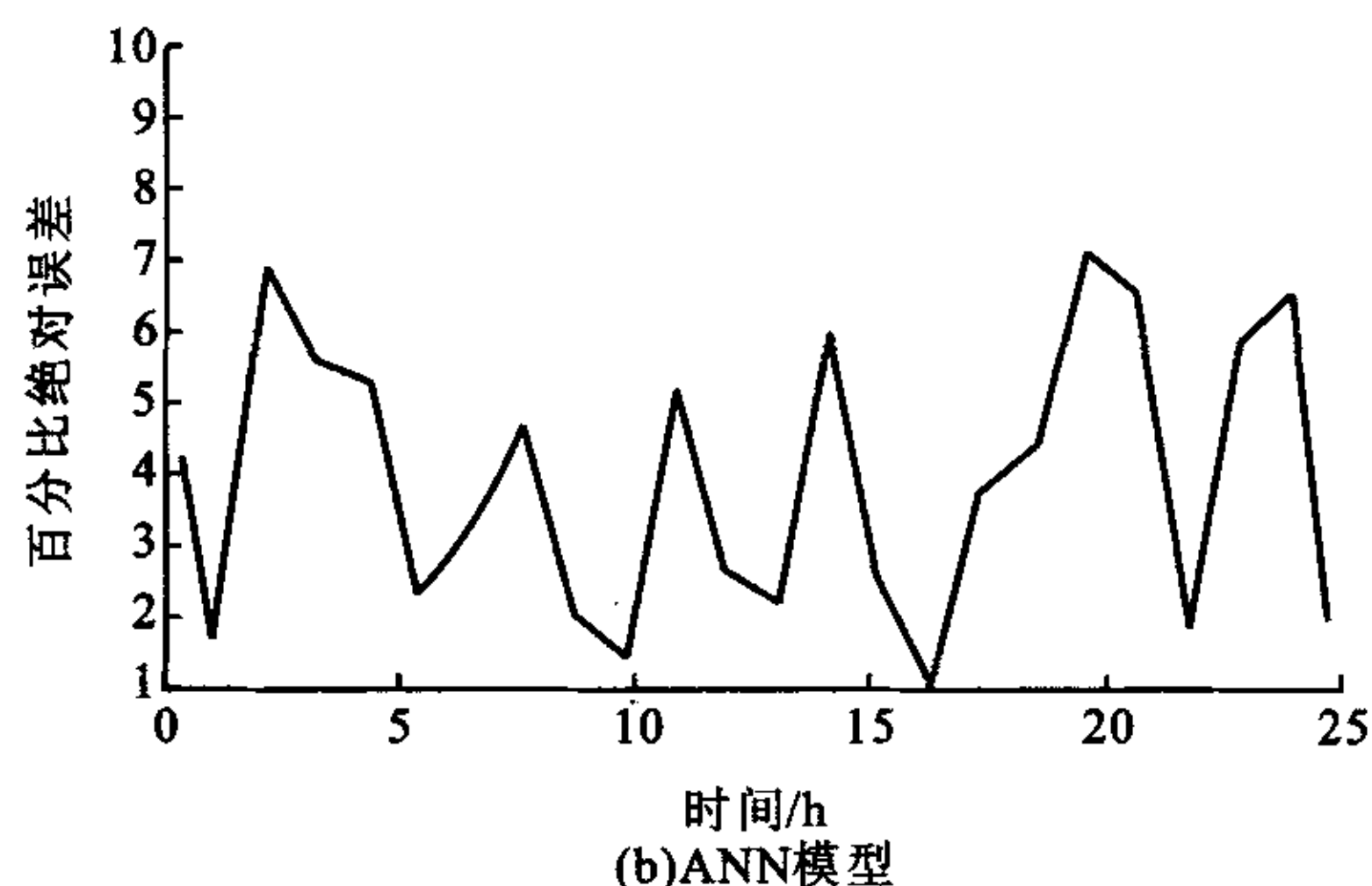
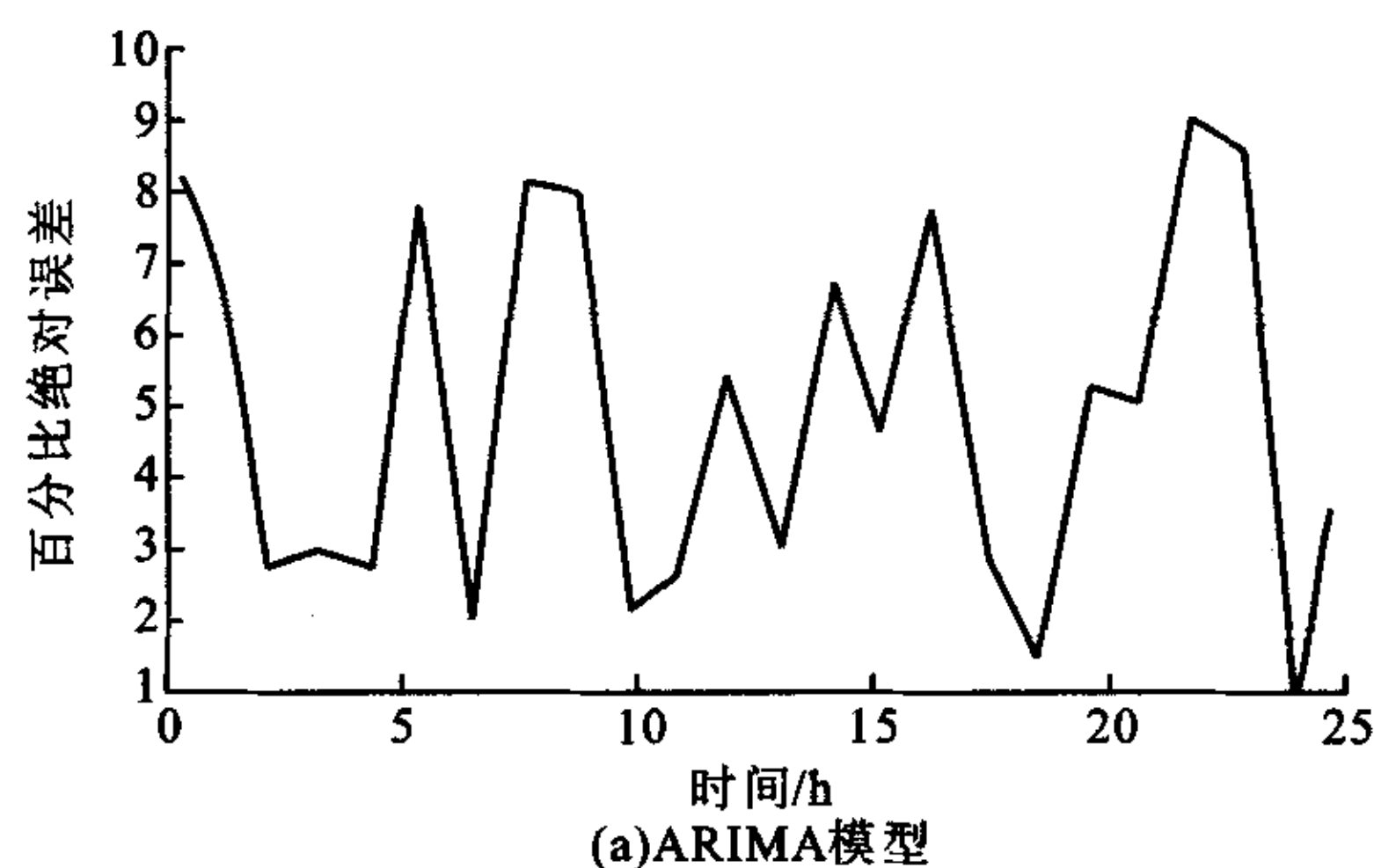


图 2 3 种方法的百分比绝对误差

Fig. 2 Percentage Absolute Errors by Three Approaches

表 1 3 种方法的平均百分比绝对误差和相关系数对比

Tab. 1 Comparison on Mean of Percentage Absolute Errors and Correlation Coefficients by Three Approaches

方法	ARIMA 模型	ANN 模型	组合模型
$E_{MPAE}$	4.912 5	3.925 7	2.274 4
$r^2$	0.991 0	0.995 7	0.998 2



能更差。

### 3 结 语

在短时交通流量预测中,由于交通流的随机性和高度非线性,单纯利用某一种预测方法,难以得到满意的预测结果。本文中根据短时交通流量的实测数据,用ARIMA与ANN组合模型对其进行预测,得到了较好的预测效果。这2种模型的组合为智能交通控制与车辆诱导提供了新的方法。由于组合模型比单一模型的计算量大,当预测时间间隔进一步缩短,如何选择更加合适的模型和组合方法,使之满足实时交通流预测的需要,是值得进一步研究的课题。

#### 参考文献:

#### References:

- [1] 史其信,郑为中. 道路网短期交通流预测方法比较[J]. 交通运输工程学报,2004,4(4):68-83.  
SHI Qi-xin, ZHENG Wei-zhong. Short-term Traffic Flow Prediction Methods Comparison of Road Networks[J]. Journal of Traffic and Transportation Engineering, 2004, 4(4): 68-83.
- [2] 韩超,宋苏,王成红. 基于ARIMA模型的短时交通流实时自适应预测系统[J]. 仿真学报,2004,16(7):1530-1535.  
HAN Chao, SONG Su, WANG Cheng-hong. A Real-time Short-term Traffic Flow Adaptive Forecasting Method Based on ARIMA Model[J]. Journal of System Simulation, 2004, 16(7): 1530-1535.
- [3] 四兵锋,孙壮志,赵小梅. 基于随机用户平衡的混合交通网络流量分离模型[J]. 中国公路学报,2006,19(1):93-98.  
SI Bing-feng, SUN Zhuang-zhi, ZHAO Xiao-mei. Mixed Traffic Network Flow-split Model Based on Stochastic User Equilibrium[J]. China Journal of Highway and Transport, 2006, 19(1): 93-98.
- [4] 熊烈强,王富,李杰. 路段交通流的动力学模型及其仿真[J]. 中国公路学报,2006,19(2):91-94.  
XIONG Lie-qiang, WANG Fu, LI Jie. Dynamical Model of Traffic Flow on Segment and Its Simulation[J]. China Journal of Highway and Transport, 2006, 19(2): 91-94.
- [5] 贺国光,李宇,马寿峰. 基于数学模型的短时交通流预测方法探讨[J]. 系统工程理论与实践,2000,20(12):51-56.  
HE Guo-guang, LI Yu, MA Shou-feng. Discussion on Short-term Traffic Flow Forecasting Methods Based on Mathematical Models[J]. Systems Engineering—Theory & Practice, 2000, 20(12): 51-56.
- [6] 杨立才,贾磊,孔庆杰,等. 粗正交小波网络及其在交通流预测中的应用[J]. 系统工程理论与实践,2005,25(8):124-129.  
YANG Li-cai, JIA Lei, KONG Qing-jie, et al. Rough Orthogonal Wavelet Network and Its Applications to the Traffic Flow Forecast[J]. Systems Engineering—Theory & Practice, 2005, 25(8): 124-129.
- [7] TAN Man-chun. Study and Implementation of a Decision Support System for Urban Transit Planning[J]. Dynamics of Continuous Discrete and Impulsive Systems Series A, 2005, 13(S1): 1737-1742.
- [8] YIN H B, WONG S C, XU J M, et al. Urban Traffic Flow Prediction Using a Fuzzy-neural Approach[J]. Transportation Research Part C, 2002, 10(2): 85-98.
- [9] SMITH B L, WILLIAMS B M, OSWALD R K. Comparison of Parametric and Nonparametric Models for Traffic Flow Forecasting[J]. Transportation Research Part C, 2002, 10(4): 303-321.
- [10] VOORT M V D, DOUGHERTY M, WATSON S. Combining Kohonen Maps with ARIMA Time Series Models to Forecast Traffic Flow[J]. Transportation Research Part C, 1996, 4(5): 307-318.
- [11] GAO Y, ER M J. NARMAX Time Series Model Prediction: Feed-forward and Recurrent Fuzzy Neural Network Approaches[J]. Fuzzy Sets and Systems, 2005, 150(2): 331-350.
- [12] BATES J M, GRANGER C W J. The Combination of Forecasts[J]. Operational Research Quarterly, 1969, 20(1): 451-468.
- [13] ZHANG G P. Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model[J]. Neurocomputing, 2003, 50(1): 159-175.
- [14] BOX G E P, JENKINS G M. Time Series Analysis, Forecasting and Control[M]. San Francisco: Holden-day, 1970.